The Coefficient of Determination in the Ridge Regression

Ainara Rodriguez Sanchez

Phd. student at University of Granada,

Campus Universitario de La Cartuja, 18071 Granada (Spain)

(e-mail: arsanchez@ugr.es)

Roman Salmeron Gómez

Quantitative methods for economics and business,

Campus Universitario de la Cartuja 18071 Granada (Spain)

University of Granada

(e-mail: romansg@ugr.es)

Catalina García García

Quantitative methods for economics and business,

Campus Universitario de la Cartuja 18071 Granada (Spain)

University of Granada

(e-mail: cbgarcia@ugr.es)

March 28, 2019

Abstract

In a linear regression, the coefficient of determination, R^2, is a relevant measure that represents the percentage of variation in the dependent variable that is explained by a set of independent variables. Thus, it measures the predictive ability of the estimated model. For an ordinary least squares (OLS) estimator, this coefficient is calculated from the decomposition of the sum of squares. However, when the model presents collinearity problems (a strong linear relation between the independent variables), the OLS estimation is unstable, and other estimation methodologies are proposed, with the ridge estimation being the most widely applied. This paper shows that the decomposition of the sum of squares is not verified in the ridge regression and proposes how the coefficient of determination should be calculated in this case.

Keywords: multicollinearity, goodness of fit, sum of squares decomposition, transformation of variables.

1 Introduction

The linear regression model is commonly used in statistical analysis to study and quantify the relationship between a dependent or explained variable ($\boldsymbol{Y}$), one or more independent or explanatory variables ($\mathbf{X}_2, \ldots, \mathbf{X}_p$) and an intercept ($\mathbf{X}_1$), establishing the following model with n **observations and** $(p-1)$ **independent variables**:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1}$$

where $\mathbf{u}$ is the random disturbance (which is supposed to be spherical), $\mathbf{X}_{n\times p}$ is a matrix of observations of the independent variables (being $\mathbf{X}_1 = (1, 1, ..., 1)^t$), and $\mathbf{Y}_{n\times 1}$ is a vector of the observations of the dependent variable.

The goodness-of-fit of model (1) is commonly measured by the coefficient of determination, R^2. According to Cornell and Berger (1987), the R^2 in the estimation with ordinary least squares (OLS) *is interpreted as the proportion of the total variation associated with the use of independent variable* $\boldsymbol{X}$*. Thus, the closer* R^2 *is to one, the greater is the proportion of the total variation in the* $\boldsymbol{Y}$ *values that is explained by introducing the independent variable* $\boldsymbol{X}$ *into the regression equation.*

Nevertheless, when there exists linear relationships or near-linear relationships among two or more explanatory variables in a linear regression model, a multicollinearity problem occurs. In this situation, OLS estimation may be unstable (Uriel et al. (1997)). As consequence of this instability, the null hypothesis of the individual significance test tends not to be rejected, while the null hypothesis is rejected in the global significance test. Moreover, the estimators' variance is very high, and coefficients are very sensitive to small changes in the data.

For this reason, alternative estimation methods are applied, including the well-known ridge estimator. This method was proposed by Hoerl and Kennard (1970b) and consists of adding a small positive **constant** on the diagonal of the matrix $\boldsymbol{X^t X}$ to solve the multicollinearity problem. The ridge estimator of $\boldsymbol{\beta}$ is given by the following expression:

$$\hat{\boldsymbol{\beta}}(k) = (\mathbf{X}^t\mathbf{X} + k\mathbf{I}_{p\times p})^{-1}\mathbf{X}^t\mathbf{Y}, \quad k \geq 0, \tag{2}$$

where $\mathbf{I}_{p\times p}$ is the identity matrix of order p.

Since $\hat{\boldsymbol{\beta}}(k) = \mathbf{Z}_k\hat{\boldsymbol{\beta}}$, with $\mathbf{Z}_k = (\mathbf{X}^t\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^t\mathbf{X}$, $\hat{\boldsymbol{\beta}}(k)$ is a biased estimator of $\boldsymbol{\beta}$ when $k > 0$ and its covariance-variance matrix is $var\left(\hat{\boldsymbol{\beta}}(k)\right) = var\left(\mathbf{Z}_k\hat{\boldsymbol{\beta}}\right) = \sigma^2 \cdot \mathbf{Z}_k(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{Z}_k$, where σ^2 is the variance of the random disturbance. When $k = 0$, the ridge estimator matches with OLS.

Although this method has been widely applied to improve the mean squared error and the numerical stability of the estimators (see Casella (1985)), it is not exempt from anomalies, as shown in the works of Smith and Campbell (1980) and Jensen and Ramirez (2008).

In this same line, the expression used for R^2 in the OLS does not apply to the ridge estimation since (as was commented in García et al. (2017) and as this work will show) the decomposition of the sum of squares verified in the OLS and used to obtain the coefficient of determination it is not fulfilled in the ridge estimation. Thus, an alternative expression for the coefficient of determination must be applied to the ridge regression. This coefficient of determination for the ridge estimation, denoted $R^2(k)$, will allow us to measure the goodness of fit of the ridge estimator.

Numerous papers treat the collinearity problem with the ridge method, but most of them do not apply the coefficient of determination. To the best of our knowledge, only McDonald (2009) and McDonald (2010) applied an expression (the square of the correlation coefficient between the dependent variable and its predicted value) that concurs with the R^2 in OLS (that is, when $k = 0$) for standardized data. Moreover, it is shown to be a strictly decreasing function in k.

The paper is structured as follows: section 2 presents the decomposition of the sum of squares for the ridge regression. Based on the traditional coefficient of determination, section 3 obtains an expression for the coefficient of determination in the ridge regression. Moreover, its monotony is studied in section 4 using a Monte Carlo simulation. Section 5 analyzes how the standardization of the dataset affects the proposed coefficient of determination. Section 6 illustrates the contribution of this paper with a numerical example. Finally, section 7 highlights the main conclusions of the paper.

2 The decomposition of the sum of squares in a ridge estimation

This section shows that the decomposition of the sum of squares that leads to the coefficient of determination in ordinary least squares (OLS) is not verified in a ridge regression.

Proposition 1 *The residuals that originate from ridge estimator,* $\mathbf{e}(k) = \mathbf{Y} - \hat{\mathbf{Y}}(k) = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}(k)$ *do not have to sum zero when* $k > 0$.

Proof 1 *On one hand, starting from a normal equations system,* $(\mathbf{X}^t\mathbf{X} + k\mathbf{I}) \cdot \hat{\boldsymbol{\beta}}(k) = \mathbf{X}^t\mathbf{Y}$, *where* $\mathbf{X}_1 = (1, 1, ..., 1)^t$, *it is had that the first row of* $\mathbf{X}^t\mathbf{X} + k\mathbf{I}$:

$$(n + k, \sum_{i=1}^{n} X_{2i}, ..., \sum_{i=1}^{n} X_{pi}),$$

multiplied by $\hat{\boldsymbol{\beta}}(k) = [\hat{\beta}_1(k), \hat{\beta}_2(k), ..., \hat{\beta}_p(k)]^t$ *it must be equal to the first element of* $\mathbf{X}^t\mathbf{Y}$, $\sum\limits_{i=1}^{n} Y_i$. *This is*

$$\sum_{i=1}^{n} Y_i = (n + k)\hat{\beta}_1(k) + \hat{\beta}_2(k) \sum_{i=1}^{n} X_{2i} + ... + \hat{\beta}_p(k) \sum_{i=1}^{n} X_{pi}.$$

On the other hand,

$$\begin{aligned} e_i(k) &= Y_i - \hat{Y}_i(k) = Y_i - (\hat{\beta}_1(k) + \hat{\beta}_2(k)X_{2i} + ... + \hat{\beta}_p(k)X_{pi}), \\ \sum_{i=1}^{n} e_i(k) &= \sum_{i=1}^{n} Y_i - (n\hat{\beta}_1(k) + \hat{\beta}_2(k) \sum_{i=1}^{n} X_{2i} + ... + \hat{\beta}_p \sum_{i=1}^{n} X_{pi}) = k \cdot \hat{\beta}_1(k), \qquad (3) \end{aligned}$$

and $k \cdot \hat{\beta}_1(k)$ *does not have to sum to zero when* $k > 0$. □

Corollary 1 *Taking into account Proposition 1, it is not verified that* $\sum\limits_{i=1}^{n} Y_i = \sum\limits_{i=1}^{n} \hat{Y}_i(k)$ *when* $k > 0$.

Proof 2 *Like* $\sum\limits_{i=1}^{n} e_i(k) \neq 0$, $\sum\limits_{i=1}^{n} Y_i \neq \sum\limits_{i=1}^{n} \hat{Y}_i(k)$ *must be true since*

$$Y_i = \hat{Y}_i(k) + e_i(k) \Rightarrow \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i(k) + \sum_{i=1}^{n} e_i(k),$$

ergo, $\bar{\mathbf{Y}} \neq \bar{\hat{\mathbf{Y}}}(k)$. □

Corollary 2 *Taking into account Proposition 1,* $\sum\limits_{i=1}^{n}(\hat{Y}_i(k) - \bar{\mathbf{Y}})e_i(k) = 0$ *is not verified when* $k > 0$*.*

Proof 3 *Taking into account that* $\sum\limits_{i=1}^{n} e_i(k) \neq 0$*:*

$$\sum_{i=1}^{n}(\hat{Y}_i(k) - \bar{\mathbf{Y}})e_i(k) = \sum_{i=1}^{n}\hat{Y}_i(k)e_i(k) - \bar{\mathbf{Y}}\sum_{i=1}^{n} e_i(k) \neq 0,$$

unless $\hat{Y}_i(k) = \bar{\mathbf{Y}}$*, which does not make sense.* □

Theorem 1 *From Corollary 2, it can be stated that the decomposition of the sum of squares in OLS is not verified in the ridge regression when* $k > 0$*.*

Proof 4 *From* $e_i(k) = Y_i - \hat{Y}_i(k)$*, it is verified that*

$$\begin{aligned} Y_i &= \hat{Y}_i(k) + e_i(k) \Rightarrow Y_i - \bar{\mathbf{Y}} = \hat{Y}_i(k) - \bar{\mathbf{Y}} + e_i(k), \\ (Y_i - \bar{\mathbf{Y}})^2 &= (\hat{Y}_i(k) - \bar{\mathbf{Y}})^2 + e_i^2(k) + 2 \cdot (\hat{Y}_i(k) - \bar{\mathbf{Y}}) \cdot e_i(k) \\ \sum_{i=1}^{n}(Y_i - \bar{\mathbf{Y}})^2 &= \sum_{i=1}^{n}(\hat{Y}_i(k) - \bar{\mathbf{Y}})^2 + \sum_{i=1}^{n} e_i^2(k) + 2 \cdot \sum_{i=1}^{n}(\hat{Y}_i(k) - \bar{\mathbf{Y}}) \cdot e_i(k). \end{aligned}$$

As $\bar{\mathbf{Y}} \neq \bar{\hat{\mathbf{Y}}}(k)$ *and* $\sum\limits_{i=1}^{n}(\hat{Y}_i(k) - \bar{\mathbf{Y}})e_i(k) \neq 0$*, the condition of the sum of squares*[1]*,* $SST = SSE + SSR$*, which is analogous to that of OLS, is not fulfilled.* □

3 Coefficient of determination in the ridge estimation

The coefficient of determination, R^2, is a useful measure in linear regressions to determine the quality of the model to replicate results and the variation rate of the results (explained variable denominated $\mathbf{Y}$) that can be explained by the model (other explanatory variables denominated $\mathbf{X}$) after the OLS estimate. In the model (1), it is defined as follows:

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum\limits_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum\limits_{i=1}^{n}(Y_i - \bar{\mathbf{Y}})^2}. \tag{4}$$

[1] Where SST is the sum of total squares, SSE is the sum of squares explained and SSR is the sum of squares of errors.

If we begin with a model with only the intercept (known like restricted model) as follows,

$$\mathbf{Y} = \beta_1 + \mathbf{v}, \tag{5}$$

where $\mathbf{v}$ is the random disturbance (which is supposed to be spherical), the sum of square residuals of the restricted model, SSR_r, will be equal to the sum of squares total, SST, of the model (1), since $SSR_r = \sum\limits_{i=1}^{n} \left(Y_i - \overline{\mathbf{Y}} \right)^2$.

Therefore, the following alternative expression for expression (4) is given:

$$R^2 = 1 - \frac{SSR}{SSR_r}. \tag{6}$$

From this expression, it is obtained that:

- If $R^2 \simeq 0$ it is had that the SSR of the model (1) is practically equal to **the coefficient of determination** of the restricted model (5), so that the inclusion of variables are unnecessary. The variables do not provide any new information since the proposed model is not adequate.

- If $R^2 \simeq 1$, the SSR of the model (1) is very small in relation to the restricted model (5). Consequently, the variables introduced provide information that makes it preferable to the restricted model.

Moreover, if model (1) has an independent term, $0 \leq R^2 \leq 1$ and the decomposition of the sum of squares is verified. Then, the expression (4) can be expressed as

$$R^2 = \frac{SSE}{SST}. \tag{7}$$

Since in the ridge estimation, the decomposition of the sum of squares is not verified (see Theorem 1), the coefficient of determination must be calculated from expression (4) or (7). Taking into account the interpretation of expression (6), we consider it appropriate to define it as

$$R^2(k) = 1 - \frac{SSR(k)}{SSR_r(k)}, \quad k \geq 0. \tag{8}$$

It is evident that this coefficient of determination depends on the parameter k, and it is verified that if $k = 0$, this coincides with the expression given by OLS.

3.1 Monotony of $R^2(k)$

This section analyzes the monotony of $R^2(k)$ with respect to k.

Proposition 2 *$SSR_r(k) = \mathbf{Y}^t\mathbf{Y} - \frac{n+2k}{(n+k)^2} \cdot n^2 \cdot \overline{\mathbf{Y}}^2$, which is increasing in k.*

Proof 5 *The ridge estimator of $\boldsymbol{\beta}$ for the restricted model (5) depends on the parameter k from the following expression $\hat{\boldsymbol{\beta}}_r(k) = (\mathbf{1}^t\mathbf{1} + k)^{-1}\mathbf{1}^t\mathbf{Y}$, $k \geq 0$. Taking into account this expression*

$$\begin{aligned}
SSR_r(k) &= \mathbf{e}_r(k)^t\mathbf{e}_r(k) = (\mathbf{Y} - \mathbf{1}\hat{\boldsymbol{\beta}}_r(k))^t(\mathbf{Y} - \mathbf{1}\hat{\boldsymbol{\beta}}_r(k)) = \mathbf{Y}^t\mathbf{Y} - 2\hat{\boldsymbol{\beta}}_r(k)^t\mathbf{1}^t\mathbf{Y} + \hat{\boldsymbol{\beta}}_r(k)^t\mathbf{1}^t\mathbf{1}\hat{\boldsymbol{\beta}}_r(k) \\
&= \mathbf{Y}^t\mathbf{Y} - 2\mathbf{Y}^t\mathbf{1}(\mathbf{1}^t\mathbf{1} + k)^{-1}\mathbf{1}^t\mathbf{Y} + \mathbf{Y}^t\mathbf{1}(\mathbf{1}^t\mathbf{1} + k)^{-1}\mathbf{1}^t\mathbf{1}(\mathbf{1}^t\mathbf{1} + k)^{-1}\mathbf{1}^t\mathbf{Y} \\
&= \mathbf{Y}^t\mathbf{Y} - 2\left(\sum_{i=1}^n Y_i\right)(n+k)^{-1}\left(\sum_{i=1}^n Y_i\right) + \left(\sum_{i=1}^n Y_i\right)(n+k)^{-1} \cdot n \cdot (n+k)^{-1}\left(\sum_{i=1}^n Y_i\right) \\
&= \mathbf{Y}^t\mathbf{Y} - 2 \cdot \frac{\left(\sum\limits_{i=1}^n Y_i\right)^2}{(n+k)} + \frac{\left(\sum\limits_{i=1}^n Y_i\right)^2}{(n+k)^2} \cdot n = \mathbf{Y}^t\mathbf{Y} + \frac{\left(\sum\limits_{i=1}^n Y_i\right)^2 \cdot (-n-2k)}{(n+k)^2} \\
&= \mathbf{Y}^t\mathbf{Y} - \frac{n+2k}{(n+k)^2} \cdot n^2 \cdot \overline{\mathbf{Y}}^2.
\end{aligned}$$

Moreover, it is fulfilled that $SSR_r(k)$ is increasing in k since $k > 0$, $n > 0$ and

$$\frac{dSSR_r(k)}{dk} = \frac{-2n^2\overline{\mathbf{Y}}^2(n+k)^2 + 2(n+k)(n+2k)n^2\overline{\mathbf{Y}}^2}{(n+k)^4} = \frac{2n^2\overline{\mathbf{Y}}^2(n+k)k}{(n+k)^4} > 0.$$

□

Proposition 3 *$SSR(k) = \mathbf{Y}^t\mathbf{Y} - 2\sum\limits_{i=1}^p \frac{\alpha_i^2}{\lambda_i+k} + \sum\limits_{i=1}^p \frac{\alpha_i^2\lambda_i}{(\lambda_i+k)^2}$, which is increasing in k.*

Proof 6 *According to McDonald (2010), we decompose $\mathbf{X}^t\mathbf{X}$ based on its eigenvalues and eigenvectors: $\mathbf{X}^t\mathbf{X} = \boldsymbol{\Gamma}\mathbf{D}_\lambda\boldsymbol{\Gamma}^t$, with $\boldsymbol{\Gamma}^t = \boldsymbol{\Gamma}^{-1}$ and $\mathbf{D}_\lambda = diag(\lambda_1, \dots, \lambda_p)$. So, $(\mathbf{X}^t\mathbf{X} + k\mathbf{I}) = \boldsymbol{\Gamma}(\mathbf{D}_\lambda + k\mathbf{I})\boldsymbol{\Gamma}^t = \boldsymbol{\Gamma}\mathbf{D}_{\lambda+k}\boldsymbol{\Gamma}^t$.*

Calling $\boldsymbol{\gamma} = \mathbf{X}^t\mathbf{Y}$ and $\boldsymbol{\alpha} = \boldsymbol{\Gamma}^t\boldsymbol{\gamma}$, then

$$\begin{aligned}
SSR(k) &= \mathbf{e}(k)^t\mathbf{e}(k) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}(k))^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}(k)) = \mathbf{Y}^t\mathbf{Y} - 2\hat{\boldsymbol{\beta}}(k)^t\mathbf{X}^t\mathbf{Y} + \hat{\boldsymbol{\beta}}(k)^t\mathbf{X}^t\mathbf{X}\hat{\boldsymbol{\beta}}(k) \\
&= \mathbf{Y}^t\mathbf{Y} - 2\mathbf{Y}^t\mathbf{X}(\mathbf{X}^t\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^t\mathbf{Y} + \mathbf{Y}^t\mathbf{X}(\mathbf{X}^t\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^t\mathbf{X}(\mathbf{X}^t\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^t\mathbf{Y} \\
&= \mathbf{Y}^t\mathbf{Y} - 2\boldsymbol{\gamma}^t\boldsymbol{\Gamma}\mathbf{D}_{\frac{1}{\lambda+k}}\boldsymbol{\Gamma}^t\boldsymbol{\gamma} + \boldsymbol{\gamma}^t\boldsymbol{\Gamma}\mathbf{D}_{\frac{1}{\lambda+k}}\boldsymbol{\Gamma}^t\boldsymbol{\Gamma}\mathbf{D}_{\lambda}\boldsymbol{\Gamma}^t\boldsymbol{\Gamma}\mathbf{D}_{\frac{1}{\lambda+k}}\boldsymbol{\Gamma}^t\boldsymbol{\gamma} \\
&= \mathbf{Y}^t\mathbf{Y} - 2\boldsymbol{\alpha}^t\mathbf{D}_{\frac{1}{\lambda+k}}\boldsymbol{\alpha} + \boldsymbol{\alpha}^t\mathbf{D}_{\frac{1}{\lambda+k}}\mathbf{D}_{\lambda}\mathbf{D}_{\frac{1}{\lambda+k}}\boldsymbol{\alpha} \\
&= \mathbf{Y}^t\mathbf{Y} - 2\sum_{i=1}^{p}\frac{\alpha_i^2}{\lambda_i + k} + \sum_{i=1}^{p}\frac{\alpha_i^2\lambda_i}{(\lambda_i + k)^2} = \mathbf{Y}^t\mathbf{Y} - \sum_{i=1}^{p}\frac{(\lambda_i + 2k)\alpha_i^2}{(\lambda_i + k)^2}.
\end{aligned}$$

Moreover, it is verified that $SSR(k)$ is increasing in k:

$$\frac{dSSR(k)}{dk} = -\frac{2\alpha_i^2(\lambda_i + k)^2 - 2(\lambda_i + 2k)\alpha_i^2(\alpha_i + k)}{(\lambda_i + k)^4} = \frac{2\alpha_i^2(\lambda_i + k)k}{(\lambda_i + k)^4} > 0,$$

*since $k > 0$ **and** $\lambda_i > 0$ by being the eigenvalues of a symmetrical matrix.* □

Corollary 3 *Taking into account Propositions 2 and 3, $\lim\limits_{k \to +\infty} SSR_r(k) = \mathbf{Y}^t\mathbf{Y} = \lim\limits_{k \to +\infty} SSR(k)$.*

Proof 7 *Immediate.* □

Then, $SSR(k)$ and $SSR_r(k)$ increase as k increases, and the monotony of $R^2(k)$ according to the expression given in (8) depends on the *rate* at which each one increases. Therefore, it is necessary **to** study this aspect more deeply.

Proposition 4 *$SSR_r(k) - SSR(k)$ is increasing in k if*

$$\frac{n^2 \cdot \overline{\mathbf{Y}}^2}{(n+k)^3} > \sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i + k)^3}. \tag{9}$$

Proof 8 *As*

$$\begin{aligned}
SSR_r(k) - SSR(k) &= 2 \cdot \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{X}^t\mathbf{Y} - \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{X}^t\mathbf{X} \cdot \widehat{\boldsymbol{\beta}}(k) - \frac{n^2 \cdot (n+2k)}{(n+k)^2} \cdot \overline{\mathbf{Y}}^2 \\
&= \sum_{i=1}^{p}\frac{\lambda_i + 2k}{(\lambda_i + k)^2}\alpha_i^2 - \frac{n^2 \cdot (n+2k)}{(n+k)^2} \cdot \overline{\mathbf{Y}}^2,
\end{aligned}$$

then

$$\begin{aligned}\frac{\partial(SSR_r(k)-SSR(k))}{\partial k} &= 2\sum_{i=1}^{p}\frac{2\cdot(\lambda_i+k)^2-(\lambda_i+2k)\cdot 2\cdot(\lambda_i+k)}{(\lambda_i+k)^4}\cdot\alpha_i^2 \\ &\quad -n^2\cdot\frac{2\cdot(n+k)^2-(n+2k)\cdot 2\cdot(n+k)}{(n+k)^4}\cdot\overline{\mathbf{Y}}^2 \\ &= \frac{2\cdot n^2\cdot k}{(n+k)^3}\cdot\overline{\mathbf{Y}}^2-\sum_{i=1}^{p}\frac{2\cdot k\cdot\alpha_i^2}{(\lambda_i+k)^3}.\end{aligned}$$

As $\frac{2\cdot n^2\cdot k}{(n+k)^3}\cdot\overline{\mathbf{Y}}^2$ *and* $\sum\limits_{i=1}^{p}\frac{2\cdot k\cdot\alpha_i^2}{(\lambda_i+k)^3}$ *are positive, it is verified that* $SSR_r(k)-SSR(k)$ *is increasing in* k *if*

$$\frac{n^2\cdot\overline{\mathbf{Y}}^2}{(n+k)^3}>\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^3}.$$

□

Theorem 2 *Taking into account Proposition 4,* $R^2(k)$ *is increasing in* k *if*

$$\frac{n^2\cdot\overline{\mathbf{Y}}^2}{(n+k)^3}>\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^3}.$$

Proof 9 *As it is verified that*

$$\frac{n^2\cdot\overline{\mathbf{Y}}^2}{(n+k)^3}>\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^3},$$

then $SSR_r(k)-SSR(k)$ *is increasing in* k*, that is, it is verified that* $SSR_r(k)>SSR(k)$ *for all* k*. Moreover, due to* $SSR_r(k)$ *and* $SSR(k)$ *are also increasing in* k*, it must be verified that* $\frac{SSR(k)}{SSR_r(k)}$ *is decreasing in* k*.*

Consequently, $R^2(k)=1-\frac{SSR(k)}{SSR_r(k)}$ *is increasing in* k *if*

$$\frac{n^2\cdot\overline{\mathbf{Y}}^2}{(n+k)^3}>\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^3}.$$

□

Theorem 3 *Taking into account Proposition 4,* $R^2(k)$ *is decreasing in* k *if*

$$\frac{n^2\cdot\overline{\mathbf{Y}}^2}{(n+k)^3}<\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^3}.$$

Proof 10 *Similar to Proof 9.* □

Taking into account expression (9), it is concluded that

- High values of n lead to low values of $\frac{n^2 \cdot \overline{\mathbf{Y}}^2}{(n+k)^3}$ and, consequently, a tendency toward a decreasing $R^2(k)$.
- High (low) values of $\overline{\mathbf{Y}}$ lead to high (low) values of $\frac{n^2 \cdot \overline{\mathbf{Y}}^2}{(n+k)^3}$ and, consequently, a tendency toward an increasing (decreasing) $R^2(k)$.

Note that an increasing or decreasing coefficient of determination ($R^2(k)$) depends on different questions being convenient for a deeper analysis with a simulation study, as presented in section 4.

3.2 Other properties of $R^2(k)$

This section develops other interesting properties of $R^2(k)$.

Corollary 4 *Taking into account Proposition 2, $SSR_r(k) \geq 0$ for all k.*

Proof 11 *As $SSR_r(k)$ is increasing in k and $SCR_r(0) = SCR_r \geq 0$, it is verified that $SSR_r(k) \geq SCR_r \geq 0$.* □

Corollary 5 *Taking into account Proposition 3, $SSR(k) \geq 0$ for all k.*

Proof 12 *As $SSR(k)$ is increasing in k and $SCR(0) = SCR \geq 0$, it is verified that $SSR(k) \geq SCR \geq 0$.* □

Corollary 6 *Taking into account Corollary 4 and 5, $R^2(k) \leq 1$ for all k.*

Proof 13 *Indeed,*

$$R^2(k) > 1 \Leftrightarrow 1 - \frac{SSR(k)}{SSR_r(k)} > 1 \Leftrightarrow \frac{SSR(k)}{SSR_r(k)} < 0,$$

which is not possible since $SSR(k) \geq 0$ and $SSR_r(k) \geq 0$. □

Corollary 7 *Taking into account Proposition 4, $R^2(k) > 0$ if*

$$\frac{n^2 \cdot \overline{\mathbf{Y}}^2}{(n+k)^3} > \sum_{i=1}^{p} \frac{\alpha_i^2}{(\lambda_i + k)^3}.$$

Proof 14 *Since it is verified that*

$$\frac{n^2 \cdot \overline{\mathbf{Y}}^2}{(n+k)^3} > \sum_{i=1}^{p} \frac{\alpha_i^2}{(\lambda_i + k)^3},$$

it is evident that $SSR_r(k) - SSR(k)$ is increasing in k, that is, it is verified that $SSR_r(k) > SSR(k)$ for all k. Then,

$$\frac{SSR(k)}{SSR_r(k)} < 1 \Leftrightarrow -\frac{SSR(k)}{SSR_r(k)} > -1 \Leftrightarrow R^2(k) = 1 - \frac{SSR(k)}{SSR_r(k)} > 0.$$

□

Corollary 8 *Taking into account Corollary 3, $\lim\limits_{k \to +\infty} R^2(k) = 0$.*

Proof 15 *Immediate since $R^2(k) = 1 - \frac{SSR(k)}{SSR_r(k)}$.* □

4 Monte Carlo simulation

As shown in Propositions 2 and 3, the $SSR_r(k)$ and $SSR(k)$ are both increasing functions in k. Consequently, it will be difficult to know at first glance the behavior of $R^2(k)$ in relation to k, as is shown in Theorems 2 and 3. For this reason, the following Monte Carlo simulation is presented.

The values are simulated (see, for example, Gibbons (1981) or Kibria (2003)) from

$$\mathbf{X}_i = \sqrt{1-\gamma^2} \cdot \mathbf{W}_i + \gamma \cdot \mathbf{W}_p,$$

where $i = 2, \dots, p$ with $p = 3, 4, 5$, $\mathbf{W}_i \sim N(30, 10)$, $\gamma \in \{0.95, 0.96, 0.97, 0.98, 0.99\}$ and $n \in \{15, 20, 25, 30, \dots, 200\}$. The matrix $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2 \dots \mathbf{X}_p]$ is constructed such that $\mathbf{X}_1$ is a vector with ones (representing the constant term in model). Then, $\mathbf{Y} = \mathbf{X} \cdot \boldsymbol{\beta} + \mathbf{u}$, is generated where $\boldsymbol{\beta}$ is simulated by $N(\mu, 0.2)$ with $\mu = 10$ (unless other value will be established) and $\mathbf{u} \sim N(0, 0.1)$.

Results are displayed in Figures 1 to 5. In these figures, the left part shows the traditional values of k ($k \in [0,\ 1]$) and the right part shows an extended range to analyze the long-term behavior of $R^2(k)$. From these figures, it is possible to conclude the following:

- From figure (a) of Figures 1, 2 and 3 for $n < 50$ and $k \in [0,\ 1]$, it is observed that the coefficient of determination decreases initially until it reaches the point of inflection from which it grows. However, in Figure (c), a decreasing behavior is observed for $n > 50$ only. Thus, these figures support the previous supposition that high values of n imply a decreasing $R^2(k)$.

- From figure (b) of Figures 1, 2 and 3 for $n < 50$ and $k \in [0,\ 10]$, it is observed that the decreasing and increasing behaviors are greater than those observed in Figure (a), while in figure (d) for $n > 50$, the same behavior as in (a) and (b) is observed. That is to say, when the value of n increases, the turning point moves to the right, outside the interval $[0,\ 1]$.

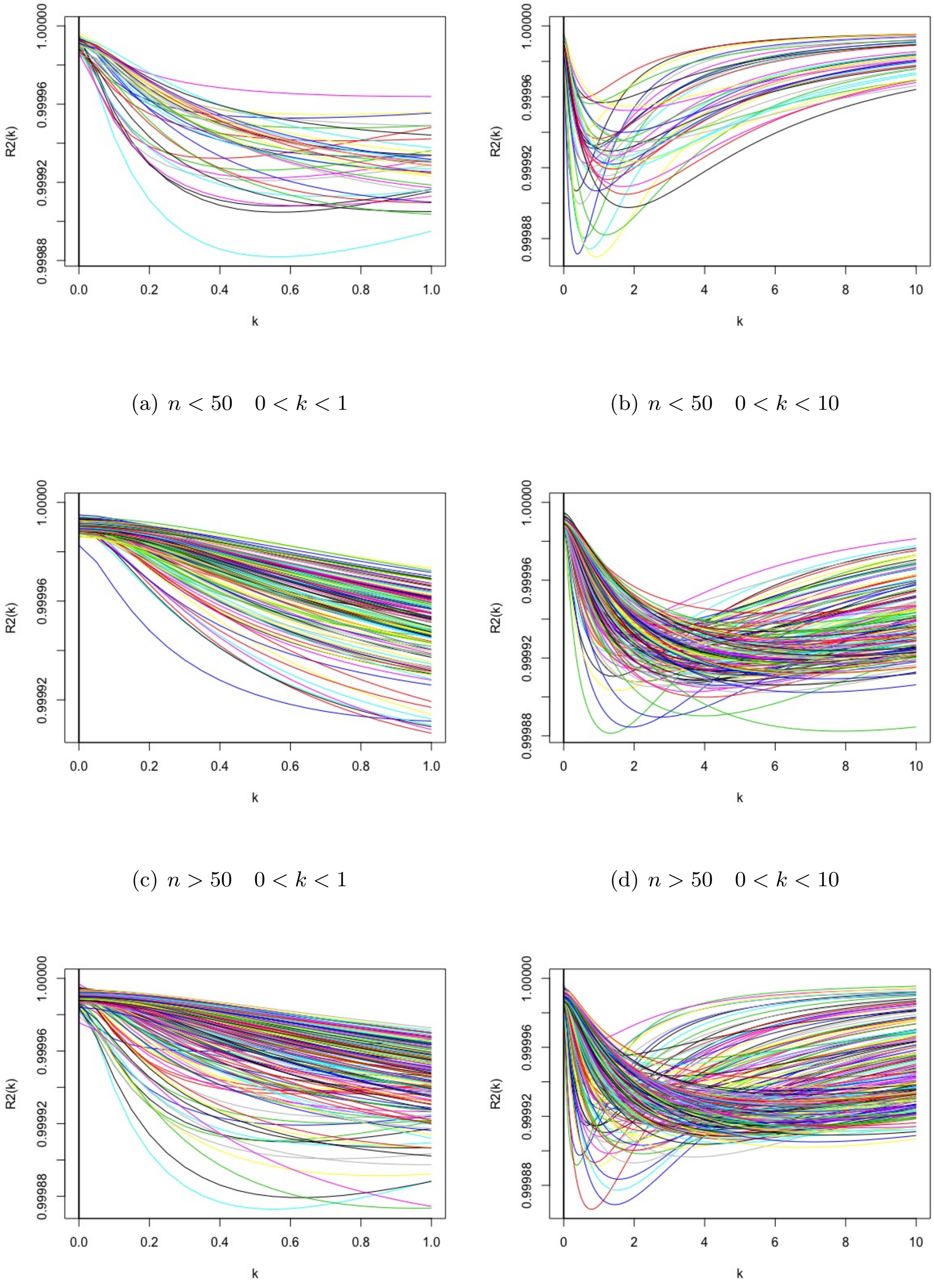
- From representations (e) and (f) of Figures 1, 2 and 3, for $n \in \{15, 20, \cdots, 200\}$, a softer behavior in simulations with $n > 50$ is observed. This representation confirms the previous comment.

- In Figure 4, it is observed that turning point (where $R^2(k)$ goes from decreasing to increasing) moves to the left when $\overline{\mathbf{Y}}$ increases. Thus, the supposition that high values of $\overline{\mathbf{Y}}$ imply an increasing $R^2(k)$ is supported for this simulation.

- In Figure 5, the asymptotic behavior shown in Corollary 8 is observed. That is to say, there is a second turning point from which the $R^2(k)$ decreases to zero.

5 Transformation of variables

The transformation of variables is a very common practice when working with an econometric model in which the collinearity is worrying. From the work of Marquandt (1980), the most applied transformation is the standardization, that is to say, to subtract their mean and divide them by the square root of n times their variance. This section analyzes how this transformation affects the coefficient of determination of the ridge regression:



(a) $n < 50 \quad 0 < k < 1$

(b) $n < 50 \quad 0 < k < 10$

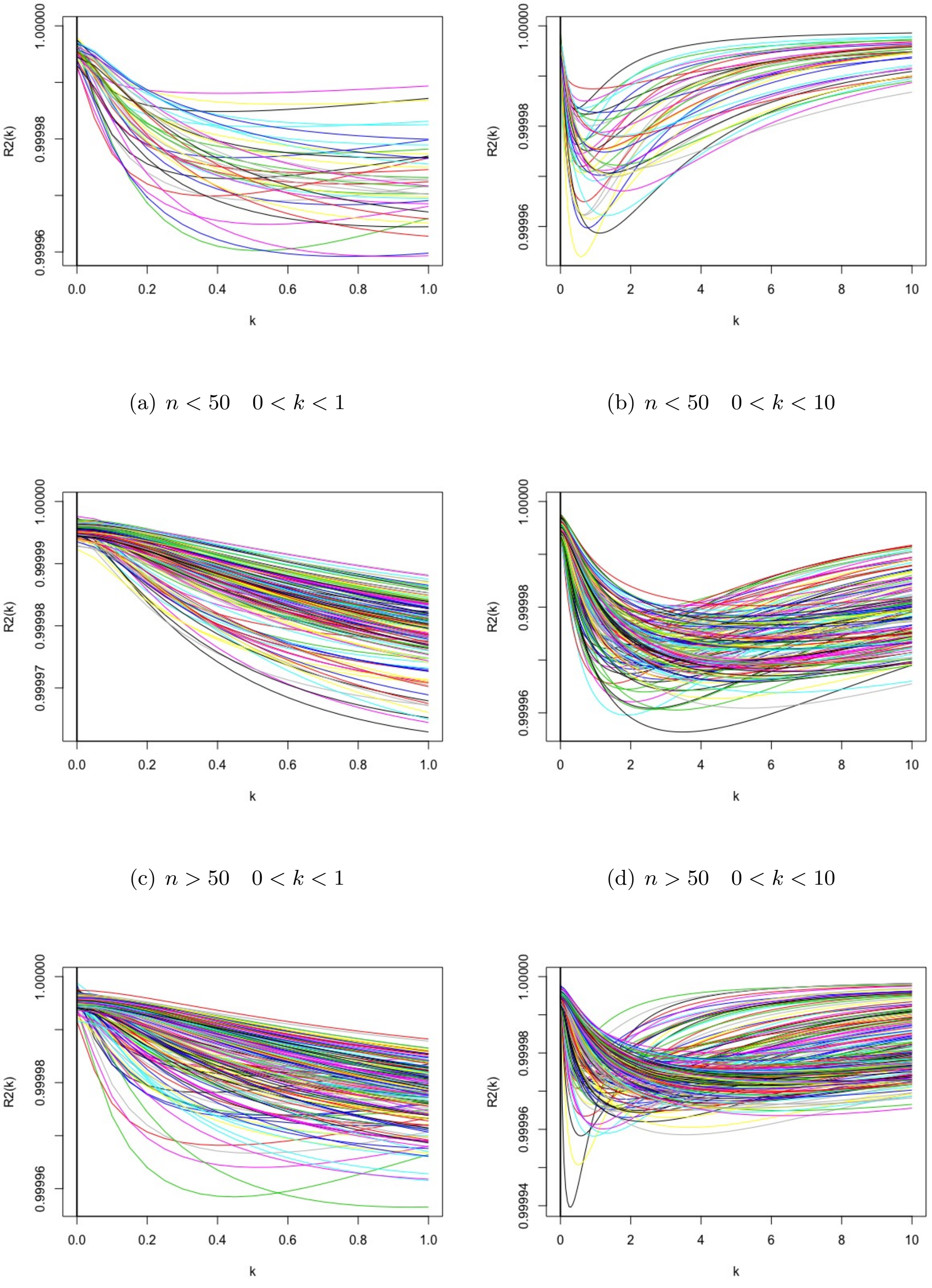
(c) $n > 50 \quad 0 < k < 1$

(d) $n > 50 \quad 0 < k < 10$

(e) $n \in \{15, 20, \cdots, 200\} \quad 0 < k < 1$

(f) $n \in \{15, 20, \cdots, 200\} \quad 0 < k < 10$

Figure 1: Simulation for $p = 3$



(a) $n < 50 \quad 0 < k < 1$

(b) $n < 50 \quad 0 < k < 10$

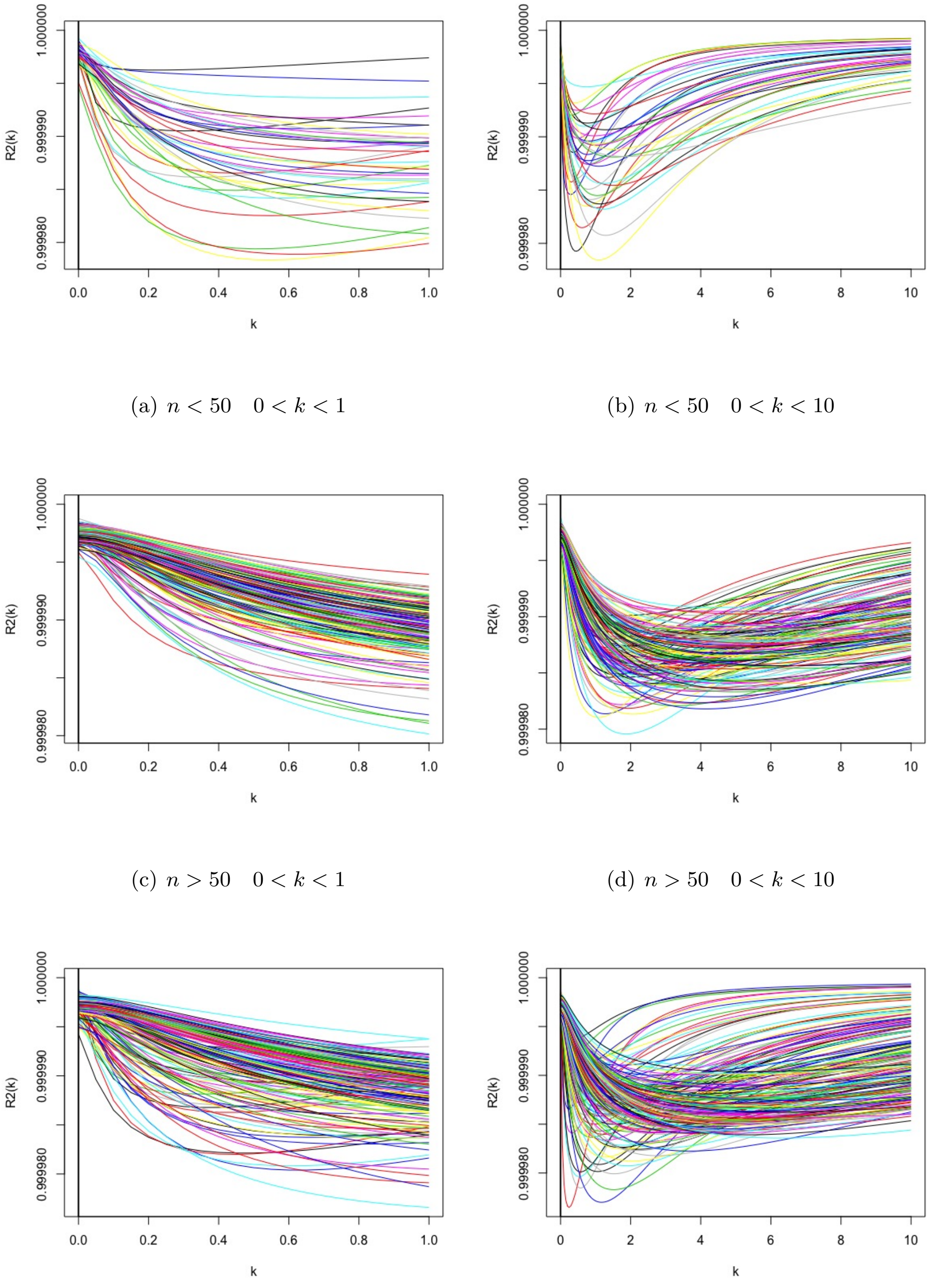
(c) $n > 50 \quad 0 < k < 1$

(d) $n > 50 \quad 0 < k < 10$

(e) $n \in \{15, 20, 25, 30, \cdots, 200\} \quad 0 < k < 1$

(f) $n \in \{15, 20, \cdots, 200\} \quad 0 < k < 10$

Figure 2: Simulation for $p = 4$



(a) $n < 50 \quad 0 < k < 1$

(b) $n < 50 \quad 0 < k < 10$

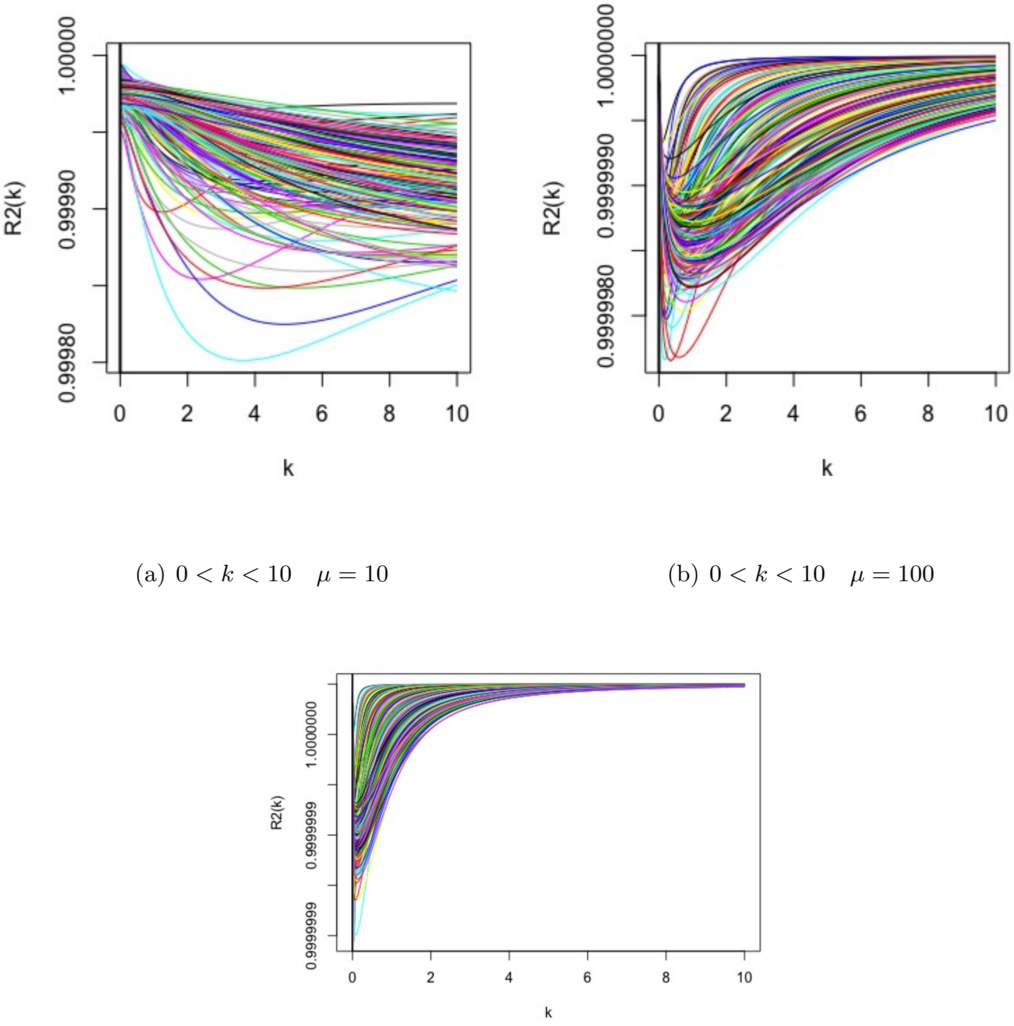
(c) $n > 50 \quad 0 < k < 1$

(d) $n > 50 \quad 0 < k < 10$

(e) $n \in \{15, 20, 25, 30, \cdots, 200\} \quad 0 < k < 1$

(f) $n \in \{15, 20, 25, 30, \cdots, 200\} \quad 0 < k < 10$

Figure 3: Simulation for $p = 5$

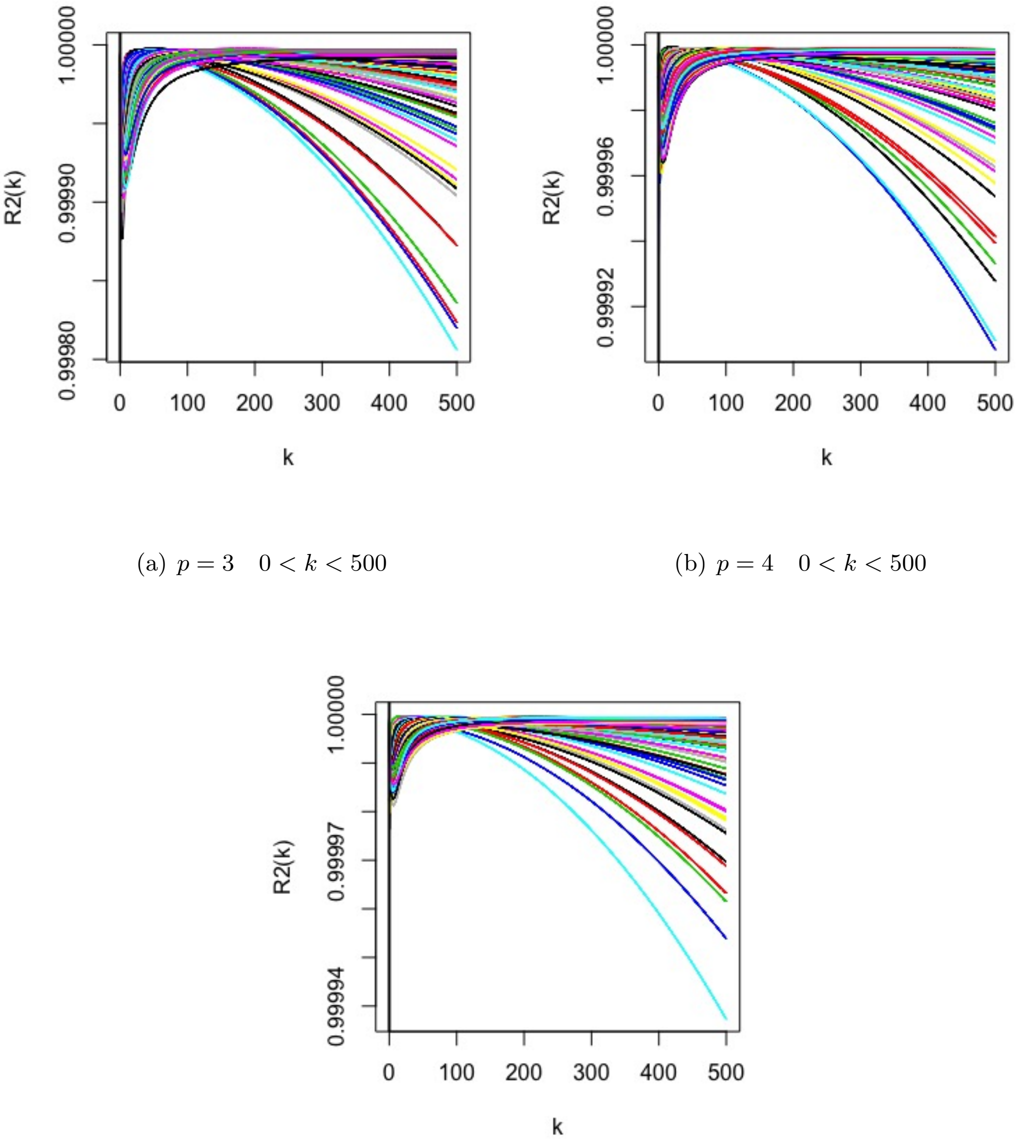


(a) $0 < k < 10 \quad \mu = 10$

(b) $0 < k < 10 \quad \mu = 100$

(c) $0 < k < 10 \quad \mu = 500$

Figure 4: Simulation for $p = 5$: $n = \{15, 20, 25, 30, \cdots, 200\}$



(a) $p = 3 \quad 0 < k < 500$

(b) $p = 4 \quad 0 < k < 500$

(c) $p = 5 \quad 0 < k < 500$

Figure 5: Simulation for $p = 3, 4, 5$, $n \in \{15, 20, 25, 30, \cdots, 200\}$ and $k \in [0,\ 500]$

Theorem 4 *Denoting* $\mathbf{y}$ *and* $\mathbf{x}$ *as the standardized versions of* $\mathbf{Y}$ *and* $\mathbf{X}$*, it is verified that*

$$R^2(k) = \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{y} + k \cdot \widehat{\boldsymbol{\beta}}(k)^t\widehat{\boldsymbol{\beta}}(k), \quad (10)$$

is decreasing in k*.*

Proof 16 *If the dependent variable is standardized, it is verified that* $\overline{\mathbf{y}} = 0$ *and* $\mathbf{y}^t\mathbf{y} = 1$*. In this case,*

$$\begin{aligned} SSR(k) &= 1 - 2 \cdot \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{y} + \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{x} \cdot \widehat{\boldsymbol{\beta}}(k) = 1 - \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{y} - k \cdot \widehat{\boldsymbol{\beta}}(k)^t\widehat{\boldsymbol{\beta}}(k), \\ SSR_r(k) &= 1, \end{aligned}$$

where it will be applied that $\widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{y} = \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{x} \cdot \widehat{\boldsymbol{\beta}}(k) + k \cdot \widehat{\boldsymbol{\beta}}(k)^t\widehat{\boldsymbol{\beta}}(k)$*. Then,*

$$R^2(k) = 1 - \frac{SSR(k)}{SSR_r(k)} = \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{y} + k \cdot \widehat{\boldsymbol{\beta}}(k)^t\widehat{\boldsymbol{\beta}}(k).$$

However, by following the steps presented in Proof 6,

$$\widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{y} = \sum_{i=1}^{p-1} \frac{\alpha_i^2}{\lambda_i + k}, \quad k \cdot \widehat{\boldsymbol{\beta}}(k)^t\widehat{\boldsymbol{\beta}}(k) = k \cdot \sum_{i=1}^{p-1} \frac{\alpha_i^2}{(\lambda_i + k)^2},$$

and then,

$$R^2(k) = \sum_{i=1}^{p-1} \frac{\lambda_i + 2 \cdot k}{(\lambda_i + k)^2} \cdot \alpha_i^2. \quad (11)$$

In this case, $R^2(k)$ *is decreasing in* k *since*

$$\frac{\partial R^2(k)}{\partial k} = -2 \cdot k \cdot \sum_{i=1}^{p-1} \frac{\alpha_i^2}{(\lambda_i + k)^3} < 0.$$

□

Corollary 9 *Taking into account Theorem 4,* $\lim\limits_{k \to +\infty} R^2(k) = 0$*.*

Proof 17 *Immediate from expression (11).* □

Thus, all problems presented in subsection 3.1 and illustrated in the simulation shown in section 4 disappear when the variables are standardized. In this case, the following coefficient of determination is obtained: a) continues in $k = 0$ since it coincides with the one given by OLS ($R^2(0) = \widehat{\boldsymbol{\beta}}(0)^t \cdot \mathbf{x}^t\mathbf{y} = \widehat{\boldsymbol{\beta}}^t \cdot \mathbf{x}^t\mathbf{y} = R^2$), b) decreasing as k increases and c) always positive.

These conclusions coincide with those obtained by García et al. (2016) and Salmerón et al. (2017), who showed that the data must be standardized to correctly calculate the variance inflation factor (VIF) and the corrected variance inflation factor (CVIF), respectively.

Finally, note that when $k = 0$, the expression given in (10) coincides with the one given in McDonald (2010) for the square of the correlation coefficient of the actual values, $\mathbf{y}$, and predicted values, $\widehat{\mathbf{y}}$, given by

$$R^2(k) = \frac{(\widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{x} \cdot \widehat{\boldsymbol{\beta}}(k) + k \cdot \widehat{\boldsymbol{\beta}}(k)^t\widehat{\boldsymbol{\beta}}(k))^2}{\widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{x} \cdot \widehat{\boldsymbol{\beta}}(k)}. \tag{12}$$

However, these values are different when $k > 0$. It is important to highlight this question since the notation used in McDonald (2010) corresponds to $R^2(k)$; thus, saying that this coefficient coincides with the coefficient of determination in OLS when $k = 0$ can lead to the wrong conclusion that the expression (12) corresponds to the coefficient of determination associated with the ridge regression.

5.1 Augmented model

Marquardt (1970) showed that the ridge estimator can be obtained from the OLS estimation of the following augmented model:

$$\mathbf{Y}_A = \mathbf{X}_A \cdot \boldsymbol{\beta}_A + \mathbf{u}_A,$$

where $\mathbf{Y}_A = \begin{pmatrix} \mathbf{Y} \\ \mathbf{0}_{p\times 1} \end{pmatrix}$ and $\mathbf{X}_A = \begin{pmatrix} \mathbf{X} \\ \sqrt{k} \cdot \mathbf{I}_{p\times p} \end{pmatrix}$ with $\mathbf{0}_{p\times 1}$ and $\mathbf{I}_{p\times p}$ a vector of zeros and the identity matrix with the adequate dimensions, respectively, due to $\widehat{\boldsymbol{\beta}}_A = \left(\mathbf{X}_A^t\mathbf{X}_A\right)^{-1} \cdot \mathbf{X}_A^t\mathbf{Y}_A = \left(\mathbf{X}^t\mathbf{X} + k \cdot \mathbf{I}\right)^{-1} \cdot \mathbf{X}^t\mathbf{Y} = \widehat{\boldsymbol{\beta}}(k)$.

In this case, the coefficient of determination will be given by

$$R_A^2(k) = 1 - \frac{\mathbf{Y}_A^t\mathbf{Y}_A - \widehat{\boldsymbol{\beta}}_A^t \cdot \mathbf{X}_A^t\mathbf{Y}_A}{\mathbf{Y}_A^t\mathbf{Y}_A - (n+p) \cdot \overline{\mathbf{Y}}_A^2} = 1 - \frac{\mathbf{Y}^t\mathbf{Y} - \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{X}^t\mathbf{Y}}{\mathbf{Y}^t\mathbf{Y} - (n+p) \cdot \left(\frac{n}{n+p}\right)^2 \cdot \overline{\mathbf{Y}}^2},$$

that it will be given by the following expression for standardized variables:

$$R_A^2(k) = 1 - \frac{1 - \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{y}}{1 - 0} = \widehat{\boldsymbol{\beta}}(k)^t \cdot \mathbf{x}^t\mathbf{y}. \tag{13}$$

Comparing expression (13) with (10), it is evident that both coincide when $k = 0$ and differ when $k > 0$. However, by following the steps given in Theorem 4 and Corollary 9, it is immediate to show that $R_A^2(k)$ is also continuous for $k = 0$, is decreasing in k and is always higher than zero.

6 Numerical example

Hoerl and Kennard (1970a) used data applied by Gorman and Toman (1970) for a regression with 10 independent variables, showing the usefulness of the ridge trace, and the squared length of the coefficient vector obtained from 15 values of k in the interval $(0,\ 1)$, concluding that stability is reached when $0.2 \leq k \leq 0.3$. Gorman and Toman (1970) used this same dataset to find the best subset of variables that avoid working with all variables.

Considering standardized data, the left part of Figure 6 presents the ridge trace previously noted, that is to say, the representation of $\widehat{\boldsymbol{\beta}}(k)$ for $k \in [0,\ 1]$. Meanwhile, the right part represents the squared length of the coefficient vector, $\widehat{\boldsymbol{\beta}}(k)^t\widehat{\boldsymbol{\beta}}(k)$, and the coefficient of determination according to expression (10) for the same values of k. In the vertical, the values of k equal to 0.2 and 0.3 have been highlighted.

Note that the coefficient of determination presents greater stability in its asymptotic behavior than the squared length of the coefficient vector. Taking into account that $R^2(0) = R^2 = 0.8966$ and $R^2(1) = 0.7389$, there is a reduction of 16.77%. In addition, Figure 7 represents the rate of change of the coefficient of determination for $k \in [0,\ 1]$ observing a change of 0.204% (the minimum value equals 0.086% and the maximum equals 0.29%) in the range $[0,\ 0.1]$, of 0.047% (the minimum value equals 0.229% and the maximum equals 0.276%) in the range $[0.1,\ 0.2]$ and of 0.026% (the minimum value equals 0.203% and the maximum equals 0.229%) in the range $[0.2,\ 0.3]$. Thus, the selection of $k = 0.25$ proposed by Hoerl and Kennard (1970a) will be justified.

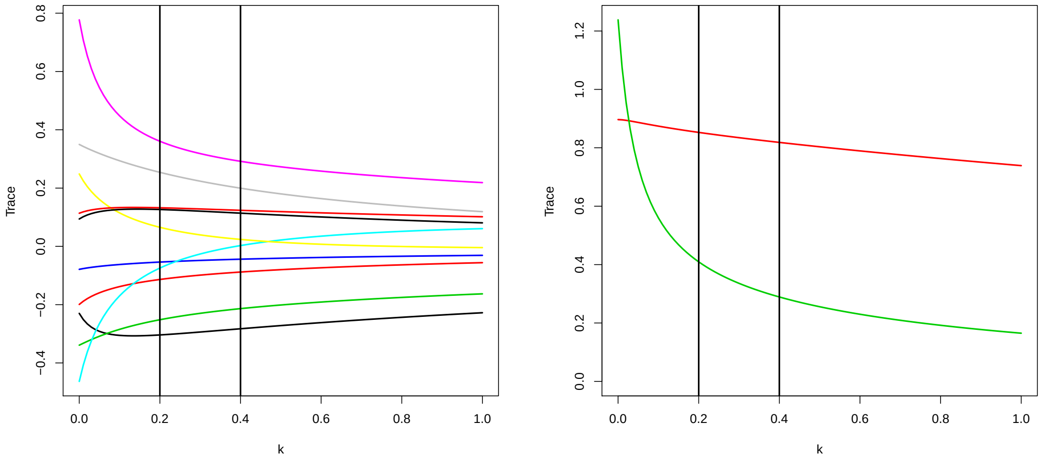


Figure 6: Graphical representation of the ridge trace (left) and the norm of the coefficients and the coefficient of determination associated with the ridge regression (right)

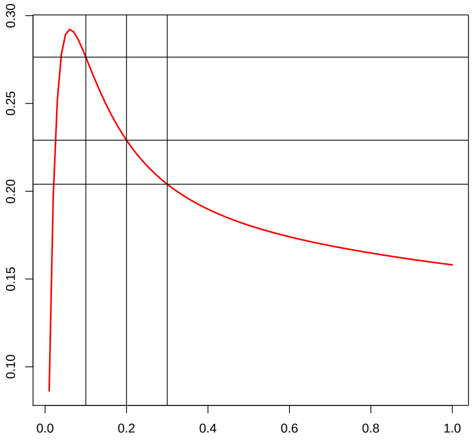


Figure 7: Graphical representation of the rate of change of the coefficient of determination

The following will be obtained:

$$\widehat{\boldsymbol{\beta}}(k) = (-0.2992,\ -0.1053,\ -0.2397,\ -0.0507,\ -0.0468,\ 0.3365,\ 0.0505,\ 0.2378,\ 0.1239,\ 0.1305)^t,$$

with a coefficient of determination of 0.8433, that is to say, for $k = 0.25$, the fit explains a 84.33% of the rate of change in the dependent variable. This figure implies a decrease of 5.33% in relation to the initial R^2 obtained by OLS.

7 Conclusion

The coefficient of determination is widely applied to analyze the goodness of fit of a linear regression estimation and how the estimated values fit the expected values. The expression of the coefficient of determination in OLS is obtained from the decomposition of the sum of squares: $SST = SSE + SSR$.

However, when a regression model presents collinearity, the OLS estimation may be unstable, and other estimation methodologies are proposed, such as the ridge regression. To the best of our knowledge, the calculation of the coefficient of determination in a ridge regression has not been treated, and this is the topic of this paper. The following conclusions are obtained:

- The decomposition of the sum of squares that is used to obtain the coefficient of determination (R^2) in OLS is not verified in a ridge regression. Consequently, the obtention of the coefficient of determination used in OLS cannot be generalized for application in a ridge regression. However, it is possible to provide a definition of the coefficient of determination from the residual sum of squares of the initial and restricted models.

- The monotony of the coefficient of determination obtained, $R^2(k)$, is analyzed with a Monte Carlo simulation that presents a different turning point (and, consequently, different decreasing/increasing behavior) depending on different factors (the size of the sample or the mean of the dependent variable). Thus, we conclude that this expression is not appropriate to measure the goodness of fit in ridge regression with original data.

- Since the standardization of the variables is a very common practice when working with econometric models with collinearity problems, we analyze how this transformation affects the definition of the coefficient of determination. We obtain that $R^2(k)$ is continuous for $k = 0$, is decreasing in k and is always higher than zero. Thus, in the same way as in García et al. (2016) and Salmerón et al. (2017), we show that the standardization in the ridge regression is not optional but compulsory to obtain the coefficient of determination and the VIF with appropriate characteristics.

- Although the application of the ridge regression mitigates the overestimation of the coefficient of the econometric model (Hoerl and Kennard (1970a)), the decrease in the coefficient of determination as the ridge factor increases indicates that the fit is worse than the OLS fit. Thus, the goodness of fit in the ridge regression for each value of k could be applied to choose an appropriate ridge factor.

- However, we showed that the coefficient of determination obtained from the augmented model presented by Marquardt (1970) differs from the coefficient of determination proposed in this paper when the data are standardized, since they coincide only when $k = 0$.

- Due to the good behavior of the standardized variable shown in section 6, the coefficient of determination could be used as a complement to the ridge trace, the squared length of the coefficient vector and other quantitative methods known to determine the adequate value of the ridge factor, k.

References

Casella, G. (1985). Conditions numbers and minimax ridge regression estimators. *Journal of the American Statistical Association 80*(391), 753–758.

Cornell, J. and R. Berger (1987). Factors that influence the value of the coefficient of determination in simple linear and nonlinear regression models. *Phytopathology 77*(1), 63–70.

García, C., R. Salmerón, A. Rodríguez, and J. García (2017). Ridge regression: some inconvenients. *Annals of Applied Economics XXXI*, 15–25.

García, J., R. Salmerón, C. García, and M. López (2016). Standardization of variables and collinearity diagnostic in ridge regression. *International Statistical Review 84*(2), 245–266.

Gibbons, D. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association 76*(373), 131–139.

Gorman, J. and R. Toman (1970). Selection of variables for fitting equations to data. *Technometrics 8*, 27–51.

Hoerl, A. and R. Kennard (1970a). Ridge regression: Applications to nonorthogonal problems. *Technometrics 12*(1), 69–82.

Hoerl, A. and R. Kennard (1970b). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*(1), 55–67.

Jensen, D. and D. Ramirez (2008). Anomalies in the foundations of ridge regression. *International Statistics Review 76*(1), 89–105.

Kibria, B. (2003). Performance of some new ridge regression estimators. *Communications in Statistics: Simulation and Computation 32*(2), 419–435.

Marquandt, D. W. (1980). You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association, Theory and Methods 75*(369), 87–91.

Marquardt, D. (1970). Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics 12*(3), 591–612.

McDonald, G. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics 1*(1), 93–100.

McDonald, G. (2010). Tracing ridge regression coefficients. *Wiley Interdisciplinary Reviews: Computational Statistics 2*(6), 695–703.

Salmerón, R., J. García, C. García, and M. López (2017). A note about the corrected vif. *Statistical Papers 58*(3), 929–945.

Smith, G. and F. Campbell (1980). A critique of some ridge regression methods. *Journal of the American Statistical Association 75*(369), 74–81.

Uriel, E., A. Periró, D. Contreras, and M. Moltó (1997). *Econometría: El Modelo Lineal.* Madrid: Ed. Alfa Centauro.